

Ann e Universitaire 2019-2020
Licence 3  conomie
Mati re: Logiciel R 4 -  conom trie Appliqu e
Nom de l'enseignant: Silvia CONCETTINI

EXAMEN

Consignes

1. Le devoir doit imp rativement  tre rendu avant les 23h59 du 22 Mars 2020 par email   l'adresse: silvia.concettini@univ-tours.fr. Aucun document ne sera accept  apr s ce d lai.
2. Chaque groupe remet un seul document contenant toutes les r ponses. Le document est soumis en format PDF/HTML et doit  tre nomm  Nom1_Nom2. Aucun autre format ou document ne sera accept .
3. Les r ponses doivent  tre num rot es comme les questions. Le code R, les figures, les r sultats des estimations et les commentaires seront inclus dans chaque r ponse.

Sujet

Le Bureau du recensement am ricain (US Census Bureau) est le premier fournisseur au monde de donn es sur les habitants et l' conomie am ricains. Les informations  conomiques et socio-d mographiques des citoyens sont utilis es de mani re extensive pour r aliser des  tudes quantitatives avec des objectifs scientifiques mais  galement pour mesurer les effets de diff rentes politiques et pour monitorer l' volution des principaux indicateurs socio- conomiques.

La base de donn es Salaire.csv recense 1084 travailleurs am ricains dans les ann es 1978 et 1985. Les variables collect es sont:

- $education_{it}$, nombre d'ann es d' tude de l'individu i dans l'ann e t
- sud_{it} , une variable dummy=1 si l'individu i vit dans le sud pendant l'ann e t et =0 sinon
- $nonblanc_i$, une variable dummy=1 si l'individu i n'est pas blanc et =0 sinon
- $femme_i$, une variable dummy=1 si l'individu i est femme et =0 sinon
- $marie_{it}$, une variable dummy=1 si l'individu i est mari /e dans l'ann e t et =0 sinon
- $experience_{it}$, nombre d'ann e d' perience de l'individu i dans l'ann e t ¹

¹L' perience est calcul e avec la formule suivante $experience = age - education - 6$.

- $syndicat_{it}$, une variable dummy=1 si l'individu i dans l'année t est inscrit au syndicat et =0 sinon
- $lsalaire_{it}$, le logarithme du salaire horaire (en dollars) de l'individu i dans l'année t
- age_{it} , l'âge (en ans) de l'individu i dans l'année t
- $anne_t$, l'année d'observation, avec $t = 1978, 1985$

Le Département du travail des États-Unis (US Department of Labor) vous commissionne une analyse économétrique avec un double objectif: d'une part on souhaite étudier la relation entre le salaire d'un individu et ses caractéristiques socio-démographiques; de l'autre on cherche à évaluer si cette relation a évolué au cours du temps et notamment entre les années 1978 et 1985.

Questions

1. Chargez la base de données et vérifiez qu'elle contient les variables indiquées.
2. Contrôlez que toutes les variables aient le bon format.
3. Obtenez un sommaire des variables quantitatives qui inclut minimum, maximum, moyenne, médiane et écart-type. Obtenez le même sommaire pour l'année 1978 et l'année 1985 séparément. Commentez.
4. Représentez les histogrammes des variables quantitatives avec la densité de probabilité sur l'axe des ordonnées. Pourquoi utiliser le logarithme du salaire au lieu du niveau de salaire? Justifiez.
5. Représentez graphiquement le logarithme du salaire en fonction des variables quantitatives (privilégiez une représentation qui différencie les observations selon l'année d'appartenance). Quelle type de relation attendez-vous entre la variable à expliquer et ces variables explicatives? Quelle intuition avez-vous quant à l'évolution de cette relation dans le temps?
6. Pour les variables qualitatives calculez et représentez graphiquement la répartition annuelle (en pourcentage) des individu dans chaque catégorie. Commentez.
7. Représentez la distribution de la variable à expliquer pour les différentes catégories des variables explicatives qualitatives. Quelle type de relation attendez-vous entre la variable à expliquer et ces variables explicatives?
8. Calculez la matrice de corrélation, représentez-la et commentez.
9. Sur la base des résultats obtenus aux questions précédentes estimez par MCO un/plusieurs modèle/s de votre choix. N'oubliez pas d'écrire l'équation de chaque modèle. Avez-vous utilisé des variables dummies? Avez-vous rajouté des variables en interaction? Avez-vous sélectionné d'autres formes fonctionnelles? Pourquoi? Interprétez en détail les résultats du modèle que vous allez retenir.
10. Y a-t-il de la colinéarité entre les variables explicatives que vous avez choisies? Réalisez vos tests et réestimez éventuellement votre modèle.
11. Les salaires sont mesurés en valeur nominale. Quels coefficients de vos estimations subiraient un changement si on mesurait les salaires dans les deux années en utilisant le prix de l'année 1978 comme référence? Vous pouvez utiliser le fait que pour toutes les observations de 1985 on a que: $\log(salaire_i/prix85) = \log(salaire_i) - \log(prix85)$ où $prix85 = 1.65$ est le coefficient de déflation.

12. La relation que vous avez estimée a-t-elle changée entre les années 1978 et 1985? Comment? Quel test pouvez vous mettre en place pour répondre à cette question?
13. Est-ce que le “gender gap” (la différence salariale entre hommes et femmes) a changé de façon significative sur les deux années?
14. Réalisez un diagnostic de votre régression. Corrigez votre modèle, le cas échéant. La variance des erreurs est-elle restée constante au cours du temps?
15. Y a-t-il des observations influentes? Y a-t-il des observations aberrantes? Réalisez vos analyses et réestimez éventuellement votre modèle.